2014

# A Demonstration of Canonical Correlation Analysis with Orthogonal Rotation to Facilitate Interpretation

Patrick V. Dattalo
*Virginia Commonwealth University*, pdattalo@vcu.edu



Recommended Citation
Dattalo, P. (2014). A demonstration of canonical correlation analysis with orthogonal rotation to facilitate interpretation. Unpublished manuscript, School of Social Work, Virginia Commonwealth University, Richmond, Virginia.

**A Demonstration of Canonical Correlation Analysis with Orthogonal Rotation to Facilitate Interpretation**

**Submitted by:**

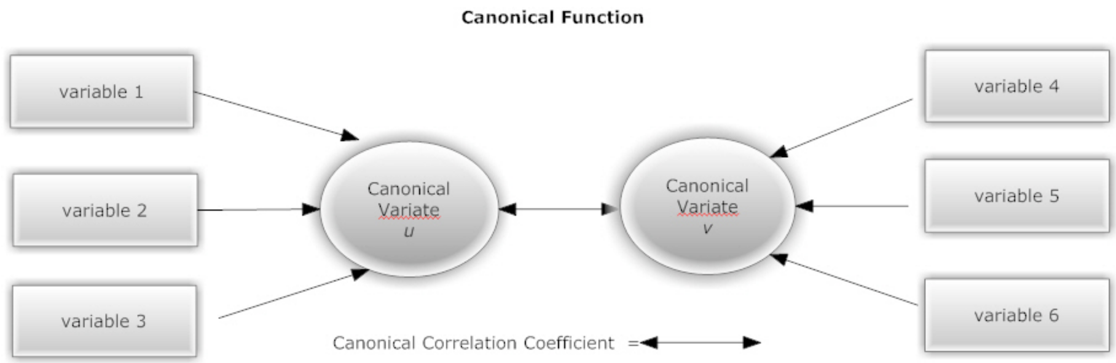
**Patrick Dattalo**
**Virginia Commonwealth University**
**School of Social Work**

## Overview and Key Terms

Originally proposed by Hotelling (1935; 1936), canonical correlation analysis (CCA) is a generalization of Karl Pearson's product moment correlation coefficient (Pearson, 1908). CCA is presented first as a general perspective on other multivariate procedures discussed in this book, including multivariate analysis of variance (MANOVA) and multivariate multiple regression (MMR) as suggested by, for example Baggaley (1981) and Thompson (1991). More specifically, Knapp (1978) demonstrated that "virtually all of the commonly encountered parametric tests of significance can be treated as special cases of canonical correlation analysis" (p. 410). Structural equation modeling (SEM), which is also discussed in this book, represents an even broader multivariate perspective, since it may incorporate measurement error estimation as part of the analysis (cf. Bagozzi, Fornell, & Larcker, 1981; Fan, 1997). CCA also is presented first because it can be used as a data reduction technique that may precede MANOVA, MMR, and SEM.

CCA models the relationships between two variable sets, with each set consisting of two or more variables. For example, let $CV_{X1} = a_1X_1 + a_2X_2 + ... + a_pX_p$ and $CV_{Y1} = b_1Y_1 + b_2 Y_2 + ... + b_m Y$, where $CV_{X1}$ and $CV_{Y1}$ are the first ***canonical variates u*** and ***v*** respectively. Together, each pair of canonical variates comprises a ***canonical function*** (see Figure 1).

Figure 1 Components of a Canonical Function




The maximum number of canonical functions that can be extracted equals the number of variables in the smallest canonical variate. For example, when the research problem involves five *u*-variables and three *v*-variables, the maximum number of canonical functions that can be extracted is three. In effect, then, CCA represents the bivariate correlation between the two canonical variates in a canonical function.

Knapp (1978) provides a detailed presentation of CCA. For Knapp (1978), familiarity with matrix algebra, including knowledge of eigenvalues and eigenvectors, is assumed. According to Knapp (1978), the first step in a CCA is the calculation of a correlation matrix of the variables in the model. A symmetric matrix of reduced rank equal to the number of variables in the smaller of the two sets is then derived from the intervariable correlation matrix, and ***canonical correlation coefficients*** ($R_c$) are quantified. More specifically, eigenvalues are computed for the matrix, with each eigenvalue equal to a squared canonical correlation coefficient. Bartlett (1948), for example, highlighted the mathematical similarities between CCA and factor analysis. Cooley and Lohnes (1971) emphasized that the canonical model selects linear functions of tests that have maximum variances, subject to the restriction of orthogonality.

A squared ***canonical correlation coefficient*** indicates the proportion of variance that the two composites derived from the two-variable sets linearly share.

Software to perform CCA analysis includes NCSS (www.ncss.com/), SAS (www.sas.com), and PASW (www.spss.com), Stata (www.stata.com). CCA will be demonstrated here with Stata. References to resources for users of PASW and SAS also are provided.

## The Canonical Correlation Analysis Procedure

The approach to CCA recommended here is as follows: (1) estimate one or more canonical functions, and calculate the magnitudes of $R_c$ and the redundancy index; (2) assess overall model fit based on the statistical significance of a multivariate $F$-test; (3) interpret the relative importance of each of the original variables the canonical functions by using standardized canonical coefficients (i.e., canonical weights) and canonical loadings (i.e., structure correlations); (4) consider the use of orthogonal rotation to facilitate interpretation of canonical functions, canonical loadings, and standardized canonical coefficients; and (5) validate the canonical correlation model.

### Estimating Canonical Functions

The first step in canonical correlation analysis is to derive one or more ***canonical functions***. Derivation of successive canonical functions is similar to the procedure used to derive a factor analysis model. That is, in factor analysis, the first factor extracted accounts for the maximum amount of variance in the set of variables, and successive factors are extracted from the residual variance of preceding factors. Accordingly, in CCA the first canonical function is derived to maximize the correlation between *u*-variables and *v*-variables. Successive functions are extracted from the residual variance of preceding functions. Since canonical functions are based on residual variance. Each function is uncorrelated (i.e., ***orthogonal***) from other functions derived from the same set of data.

The strength of the relationship between the pairs of variates is reflected by $R_c$. No generally accepted guidelines have been established regarding suitable sizes for canonical correlations. It seems logical that the guidelines suggested for significant factor loadings in factor analysis might be useful with canonical correlations, particularly when one considers that canonical correlations refer to the variance explained in the canonical variates (i.e., linear composites), not the original variables. A relatively strong canonical correlation (> 0.30, corresponding to about 10% of variance explained) may be obtained between two linear composites (i.e., canonical variates), even though these linear composites may not extract significant portions of variance from their respective sets of variables.

When squared, $R_c$ represents the amount of variance in one optimally weighted canonical variate accounted for by the other optimally weighted canonical variate. This shared variance between the two canonical variates is also termed ***canonical root*** or ***eigenvalue***. Although $R_c$ appears to be a is a simple and appealing measure of the shared variance, it may lead to some misinterpretation because the squared canonical correlation represents the variance shared by the linear composites of the sets of variables, and not the variance extracted from the sets of variables themselves.

One alternative or supplemental strategy for interpreting $R_c$ is the **redundancy index** (***$R_d$***). The redundancy index is similar to multiple regression's $R^2$ statistic. In multiple regression, $R^2$ represents the amount of variance in the dependent variable explained by the model's independent variables. Analogously, in CCA, $R_d$ is the amount of variance in the original variables of one set of variables in a canonical function that is explained by the canonical variate of the other set of variables in that canonical function. An $R_d$ can be computed for both the *u*-variable and the *v*-variable canonical variates in each canonical function. For example, an $R_d$ for the *v*-variables canonical variate represents the amount of variance in the original set of *u*-variables explained by the *v*-variables canonical variate. High redundancy suggests a high ability to predict. When there is a clearly defined relationship between IVs and DVs, a researcher will be interested primarily in the $R_d$ of the independent canonical variate in predicting the variance in the set of original variables in the dependent set. Although there also will be a $R_d$ for the dependent variate predicting the variance in the independent variables, the latter $R_d$ may not reported).

Calculating an $R_d$ is a three step process:

1. Calculate the amount of shared variance (SV) in a variable set measured by its canonical variate**.** Shared variance (SV) equals the average of the squared canonical loading. A Canonical loading measures the simple linear correlation between an original observed variable in the *u*- or *v*-variable set and that set's canonical variate. Canonical loadings are discussed further below in the section entitled "Interpreting the Canonical Variates."

2. Calculate the amount of shared variance between the *u* and the *v* canonical variates; namely, the canonical root. This is, calculate $R^2$; and

3. The redundancy index of a variate is then derived by multiplying the two components (shared variance of the variate multiplied by the squared canonical correlation) to find the amount of shared variance explained by the opposite variate.

To have a high redundancy index, one must have a high canonical correlation and a high degree of shared variance explained by its own variate. A high canonical correlation alone does not ensure a valuable canonical function. Redundancy indices are calculated for both the dependent and the independent variates, although in most instances the researcher is concerned only with the variance extracted from the dependent variable set, which provides a much more realistic measure of the predictive ability of canonical relationships. The researcher should note that although the canonical correlation is the same for both variates in the canonical function, the redundancy index will most likely vary between the two variates, because each will have a different amount of shared variance:

$$R_d = \mathrm{SV} * \mathrm{R_c}^2$$

$R_d$ of a canonical variate, then, is shared variance explained by its own set of variables multiplied by the squared canonical correlation ($R_c^2$) for the pair of variates. To have a high $R_d$, one must have a high canonical correlation and a high degree of shared variance explained by the dependent variate. A high canonical correlation alone does not ensure a valuable canonical function. The $R_d$ can only equal one when the synthetic variables for the function represent all the variance of every variable in the set, and the squared $\mathrm{R_c}$ also equals one.

A test for the significance of $R_d$ has been proposed by Cleroux and Lazraq (2002), but has not been widely utilized. Takane and Hwang (2005) have criticized Cleroux and Lazraq's proposed test as ill conceived. A major problem is that it regards each redundancy component as if it were a single observed predictor variable, which cannot be justified except for the rare situations in which there is only one predictor variable. Consequently, the proposed test may leads to biased results, particularly when the number of predictor variables is large, and it cannot be recommended for use. This is shown both theoretically and by Monte Carlo studies.

In summary, canonical correlation reflects the percent of variance in the dependent canonical variable explained by the independent canonical variable and is used when exploring relationships between the independent and the dependent set of variables. In contrast, redundancy has to do with the percent of variance in the set of original individual dependent variables explained by the independent canonical variable and is used when assessing the effectiveness of the canonical analysis in capturing the variance of the original variables. It is important to note that, although the canonical correlation is the same for both variates in the canonical function, $R_d$ may vary between the two variates (Hair, et al., 1998). That is, as each variate will have a differing amount of shared variance. As with the $R_c$, the researcher must determine whether each redundancy index is sufficiently large to justify interpretation in light of its theoretical and practical significance to the research problem being investigated to determine. Because CCA optimizes $R_c$, Cramer and Nicewander (1979) argue that redundancy coefficients are not truly multivariate, and that it is contradictory to calculate and interpret an $R_d$ as part of a CC. However, it is suggested here that, at a minimum, $R_d$ should be considered as an additional perspective on the meaning of an $R_c$. That is, $R_d$ may help to assess the practical significance of $R_c$. With large sample sizes, a relatively small $R_c$ (e.g., < .30) may achieve statistical significance. For example, $R_c$ explains 9 percent of the variance in a relationship between two sets of variables). Calculating an $R_d$ may allow a researcher to maintain a perspective on actual variance being explained by a canonical root: How much of the variability in one set of variables is explained by the other.

## Assessing Overall Model Fit

Usual practice is to analyze functions whose canonical correlation coefficients are statistically significant (e.g., $p < 0.05$). Multivariate test of the statistical significance of all canonical roots include ***Wilks' lambda***, ***Hotelling's trace***, ***Pillai's trace***, and ***Roy's largest root***. Before discussing these tests, three concepts from matrix algebra are briefly defined: eigenvalue, and determinant. These abstract mathematical concepts, which provide information about a matrix, may be most easily understood through examples. Readers interested in a more detailed discussion of these concepts than will be provided here should consult, for example, Carroll and Green (1997).

An ***m × n matrix*** is a rectangular array of real numbers with m rows and n columns. Rows are horizontal and columns are vertical. The ***determinant of a matrix*** is a summary measure of the *total* variance in that matrix when intercorrelations among variables are taken into account. Synonymously, a determinant is a measure of the area (or volume) of a shape of a matrix defined by its rows and columns. The 2×2 matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

has a determinant, defined as $|A| = ad - bc$.

An ***eigenvalue*** provides quantitative information about the variance in a *portion* of a data matrix. Specifically, if A is a linear transformation represented by a matrix **A** such that $\mathbf{A}X = \lambda X$ for some scalar $\lambda$, then $\lambda$ is called the eigenvalue of **A** with corresponding eigenvector **X**. In the context of factor analysis, an eigenvalue is a constant value that is associated with a factor in the analysis. An eigenvalue of 1 associated with a factor indicates that the factor explains an amount of variance equal to the amount of variance explained by an average variable in the model.

Perhaps the most widely used multivariate F-test is ***Wilk's Lamda* (Λ)**, where

$$\Lambda = \left|\frac{W}{T}\right| = \frac{|W|}{|B+W|} 0 \leq \Lambda \leq 1$$

$|W|$ and $|T|$ are the determinants of the within and total sum of squares cross-products matrices. ***W*** is the within group variability, where each score is deviated about its group mean for each variable. ***T*** is total variability, where each score is deviated about the grand mean for each variable. ***B*** is the between group variability. Within the context of CCA, ***B*** is a measure of the differential relationship of one set (e.g., the IVs) with another set (e.g., the DVs). Wilk's *Λ* is an inverse criterion: the smaller the value of L, the more evidence for the relationship of the IVs with the DVs. If there is no association between the two sets of variables, the ***B*** = 0 and ***Λ*** = $\frac{|W|}{|0+W|}$ = 1; if ***B*** were very large relative to ***W*** then ***Λ*** would approach 0.

Wilk's Λ also can be expressed as a product of the eigenvalues of $\frac{W}{T}$ expressed in matrix notation as $\boldsymbol{WT^{-1}}$. The other aforementioned three multivariate *F*-tests also can be expressed as a function of eigenvalues as follows: Roy's largest root equals the largest eigenvalue of $\boldsymbol{BW^{-1}}$, Hotelling- lawley trace equals the sum of the eigenvalues of $\boldsymbol{BW^{-1}}$, and Pillai-Bartlett trace equals the sum of the eigenvalues of $\boldsymbol{BW^{-1}}$.

In terms of power, none of the aforementioned *F*-tests is always the choice with the greatest statistical power. The Pillai-Bartlett trace is considered the most robust to violations of assumptions, Wilk's is the most widely used and consequently more likely to be familiar to readers (Warner, 2008). The Pillai-Bartlett trace is the most conservative of these four *F*-tests, but is a viable alternative if there are reasons to suspect that the assumptions of CCA are untenable.

## Interpreting the Canonical Variates

If, based on a multivariate *F*-test, (1) the canonical relationship is statistically significant, and (2) the magnitudes of $R_c$ and the redundancy index seem to suggest practical significance, the researcher may then interpret the relative importance of each of the original variables in the canonical functions. Methods for interpreting the relative importance of each of the original variables include (1) standardized canonical coefficients (i.e., canonical weights); and (2) canonical loadings (i.e., structure correlations).

The traditional approach to interpreting canonical functions involves examining the sign and the magnitude of the ***standardized canonical coefficients*** assigned to each variable in its canonical variate. See chapter 4 for a more detailed discussion of standardized coefficients within an analogous regression context. Variables with relatively larger standardized canonical coefficients contribute more to the variates. Similarly, variables whose standardized canonical coefficients have opposite signs exhibit an inverse relationship with each other, and variables with standardized canonical coefficients of the same sign exhibit a direct relationship. However, interpreting the relative importance or contribution of a variable by its canonical weight is subject to the same criticisms associated with the interpretation of beta weights in regression techniques. For example, a small weight may mean either that its corresponding variable is irrelevant in determining a relationship, or that it has been partialed out of the relationship because of high degree of multicollinearity. Another problem with the use of canonical weights is that these weights are subject to considerable instability (variability) from one sample to another. This instability occurs because the computational procedure for canonical analysis yields weights that maximize the canonical correlations for a particular sample of observed dependent and independent variable sets. These problems suggest caution in using standardized canonical coefficients to interpret the results of a canonical analysis.

Canonical loadings have been increasingly used as a basis for interpretation because of the deficiencies inherent in canonical weights. ***Canonical loadings***, also called structure coefficients, measure the simple linear correlation between an original observed variable in the *u*- or *v*-variable set and that set's canonical variate. The canonical loading reflects the variance that the observed variable shares with the canonical variate and can be interpreted like a factor loading in assessing the relative contribution of each variable to each canonical function. The methodology considers each independent canonical function separately and computes the within-set variable-to-variate correlation. The larger the coefficient, the more important it is in deriving the canonical variate. Also, the criteria for determining the significance of canonical structure correlations are the same as with factor loadings in factor analysis (e.g., 0.30, 0.50, and 0.70 are frequently used thresholds for considering a loading practically significant).

Canonical loadings, like weights, may be subject to considerable variability from one sample to another. This variability suggests that loadings, and hence the relationships ascribed to them, may be sample-specific, resulting from chance or extraneous factors. Although canonical loadings are considered relatively more valid than weights as a means of interpreting the nature of canonical relationships, the researcher still must be cautious when using loadings for interpreting canonical relationships, particularly with regard to the external validity of the findings.

Each of the aforementioned two methods for interpreting canonical variates (standardized canonical coefficients and canonical loadings) provides a unique perspective on the variates. Researchers should consider utilizing both methods. If the results of the two methods converge, there is evidence for the veracity of these results. If the results are inconsistent, then the researcher has an opportunity to further explore the relationships between and among the variables in the model being analyzed.

**Rotation of Structure and Canonical Coefficients**

There are similarities between principal components factor analysis (PCFA) and CCA. Both are variable reduction schemes that use uncorrelated linear combinations. In PCFA, generally the first few linear combinations (the components) account for most of the total variance in the original set of variables, whereas in CCA the first few pairs of linear combinations (canonical variates) generally account for most of the between association. Also, interpreting the principal components, we used the correlations between the original variables and the canonical variates will again be used to name the canonical variates.

It has been argued that often the interpretation of the components can be difficult, and that a rotation (e.g., Varimax) may be quite useful in obtaining factors that tend to load high on a small number of variables (Finn, 1978). Only the canonical covariates corresponding to significant canonical correlations should be rotated, in order to ensure that the rotated variates still correspond to significant association. The situation, however, is more complex, since two sets of factors (the successive pairs of canonical covariates) are being simultaneously rotated. Cliff and Krus (1976) showed mathematically that such a procedure is sound; the practical implementation of the procedure is possible (Finn, 1978). Cliff and Krus (1976) also demonstrated, through an example, how interpretation is made clearer through rotation.

Other researchers (c.f. Rencher, 1992) do not recommend rotation of the canonical variate coefficients. When a pair of canonical variate coefficients (i.e., *v*- and *u*-variables) is rotated, variance will be spread more evenly across the pair, and CCA's maximization property is lost. Consequently, researchers must decide if they are willing to sacrifice maximization for increased interpretability.

## Model Validation

In the last stage of CCA, the model should be validated. If sample size permits, one approach to validation is sample splitting, which involves creating two subsamples of the data and performing a CCA analysis on each subsample. Then, the results can be compared. Differences in results between subsamples suggest that these results may not generalize to the population.

## Sample Size Requirements

Stevens (1996) provides a thorough discussion of the sample size for CCA. To estimate the canonical loadings, only for the most important canonical function, Stevens recommends a sample size at least 20 times the number of variables in the analysis. To arrive at reliable estimates for two canonical functions, a sample size of at least 40 to 60 times the number of variables in the analysis is recommended.

Another perspective on estimating sample size for CCA is provided by Barcikowski and Stevens (1975). These authors suggest that CCA may detect stronger canonical correlations (e.g.,

$R > 0.7$), even with relatively small samples (e.g., $n = 50$). Weaker canonical correlations (e.g., $R = 0.3$) require larger sample sizes ($n > 200$) to be detected. Researchers should consider combining both perspectives to triangular on a minimally sufficient sample size for CCA. That is, they should consider the number of canonical functions to be interpreted, and the relative strength of the canonical loadings of the variables represented by the functions of interest.

## Strengths and Limitations of CCA

Important limitations of CCA are as follows: (1) $R_c$ reflects the variance shared by the linear composites of the sets of variables, and not the variance extracted from the variables; (2) $R_c$ is derived to maximize the correlation between linear composites, not to maximize the variance extracted; and (3) it may be difficult to identify meaningful relationships between the subsets of *u*- and *v*-variables. That is, procedures that maximize the correlation do not necessarily maximize interpretation of the pairs of canonical variates; therefore canonical solutions are not easily interpretable. $\mathrm{R_c}$ may be high, but the $\mathrm{R_d}$ maybe low.

CCA, however, can provide and effective tool for gaining insight into what otherwise may be an unmanageable number of bivariate correlations between sets of variables. CCA is a descriptive technique which can define structure in both the dependent and independent variates simultaneously. Therefore, situations where a series of measures are used for both dependent and independent variates are a logical choice for application of CCA. Canonical correlation also has the ability to define structure in each variate (i.e., multiple variates representing orthogonal functions) which are derived to maximize their correlation. Accordingly, the approach recommended here is to the use of, at least, the following four criteria to decide which canonical functions should be interpreted: (1) level of statistical significance based on a multivariate *F*-test of all canonical functions; (2) level of statistical significance of each function; (3) magnitude of the canonical correlation; and (4) redundancy measure for the percentage of variance accounted for from the two data sets.

## Annotated Example

A study is conducted to examine the relationship between factors that influence post-adoption service utilization and positive adoption outcomes. Specifically, the study tests a model that links (1) factors influencing the utilization of post-adoption services (*parents' perceptions of self-efficacy*, *relationship satisfaction between parents*, and *attitudes toward adoption*) with (2) *service utilization*, and (3) *positive adoption outcomes* (satisfaction with parenting and satisfaction with adoption agency).

The researcher performs a canonical correlation analysis as follows (all Stata commands are numbered in sequence and highlighted in bold italics):

***Variables in the First Set (i.e., the u-variables)***

- Parents' perceptions of self-efficacy (scale score)
- Relationship satisfaction between parents (scale score)
- Attitudes toward adoption (scale score)

***Variables in the Second Set (i.e., the v-variables)***

- Service utilization (the number of times client contacted agency since becoming a client)

- Satisfaction with parenting (scale score)
- Satisfaction with adoption agency (scale score)

As an orientation to the relationships between pairs of variables in these data, Figure 2 displays the bivariate correlation matrix of the six variables in the model.

Figure 2

```
. pwcorr

             | self_e~y relati~t knowle~s attitu~n servic~n satisf~g satisf~y
-------------+---------------------------------------------------------------
self_effic~y |   1.0000
relationsh~t |   0.4456   1.0000
knowledge_~s |   0.0620   0.0625   1.0000
attitude_a~n |   0.0121   0.0867   0.5385   1.0000
service_ut~n  |  -0.4491  -0.5281  -0.0974  -0.0663   1.0000
satisfacti~g |  -0.0570   0.0201   0.0122   0.7983  -0.0042   1.0000
satisfacti~y |  -0.0031  -0.1298  -0.0385  -0.0425   0.1371  -0.0187   1.0000
```

**Estimation of Canonical Functions**

**1.** ***canon (self_efficacy relationship_sat attitude_adoption) (service_utilization satisfaction_parenting satisfaction_adopt_agency), test (1 2 3)***

The first part of the output for canonical correlation analysis consists of (1) the raw canonical coefficients, (2) standard errors, (3) Wald *t*-tests, (4) *p*-values, (5) confidence intervals, and (6) the canonical correlation coefficient for each function. Note that 2, 3, 4, 5 are for the raw coefficients (see Figure 3).

Figure 3

Linear combinations for canonical correlations Number of obs = 300

| | (1) Coef. | (2) Std. Err. | (3) t | (4) P>\|t\| | (5) [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| u1 | | | | | | |
| self_effic~y | -.0111615 | .0091079 | -1.23 | 0.221 | -.0290851 | .0067621 |
| relationsh~t | -.0012394 | .0100057 | -0.12 | 0.901 | -.02093 | .0184511 |
| attitude_a~n | .2041443 | .0088799 | 22.99 | 0.000 | .1866692 | .2216194 |
| v1 | | | | | | |
| service_ut~n | -.0746464 | .0872575 | -0.86 | 0.393 | -.2463631 | .0970703 |
| satisfacti~g | .260643 | .0113056 | 23.05 | 0.000 | .2383943 | .2828917 |
| satisfacti~y | -.0057871 | .0090409 | -0.64 | 0.523 | -.0235789 | .0120047 |
| u2 | | | | | | |
| self_effic~y | -.086844 | .0171857 | -5.05 | 0.000 | -.1206642 | -.0530238 |
| relationsh~t | -.1455405 | .0188798 | -7.71 | 0.000 | -.1826947 | -.1083863 |
| attitude_a~n | .0031284 | .0167556 | 0.19 | 0.852 | -.0298455 | .0361023 |
| v2 | | | | | | |
| service_ut~n | 1.98803 | .1646469 | 12.07 | 0.000 | 1.664016 | 2.312043 |
| satisfacti~g | .0121457 | .0213327 | 0.57 | 0.570 | -.0298355 | .054127 |
| satisfacti~y | .0049308 | .0170593 | 0.29 | 0.773 | -.0286406 | .0385022 |
| u3 | | | | | | |
| self_effic~y | .1913321 | .1132978 | 1.69 | 0.092 | -.03163 | .4142943 |
| relationsh~t | -.1795836 | .1244667 | -1.44 | 0.150 | -.4245253 | .0653582 |
| attitude_a~n | .0200406 | .1104626 | 0.18 | 0.856 | -.1973422 | .2374233 |
| v3 | | | | | | |
| service_ut~n | -.325447 | 1.085447 | -0.30 | 0.765 | -2.46153 | 1.810636 |
| satisfacti~g | .0117319 | .1406373 | 0.08 | 0.934 | -.2650325 | .2884962 |
| satisfacti~y | .2087288 | .1124644 | 1.86 | 0.064 | -.0125934 | .4300509 |

(6) (Standard errors estimated conditionally)

Canonical correlations:
0.8020 0.5798 0.1073

This first part of the output is further divided into one section for each of the canonical functions; in this case there are three functions because the number of canonical functions is equal to the number of variables in the smaller of the *u*- and *v*-variable sets. That is, the *u*-variables include *parents' perceptions of self-efficacy*, *relationship satisfaction between parents*, and *attitudes toward adoption*; and the *v*-variables include *service utilization*, *satisfaction with parenting*, and *satisfaction with adoption agency*. The standard error of each test is calculated as the average conditional standard error across all students.

The unstandardized or "raw"canonical coefficients are the weights of the *u*-variables and the *v*-variables, which maximize the correlation between the two sets of variables. That is, the unstandardized canonical coefficients indicate how much each variable in each set is weighted to create the linear combinations that maximize the correlation between the two sets. The unstandardized canonical coefficients are interpreted in a manner analogous to interpreting unstandardized regression coefficients. For example, for the variable parents' perceptions of self-efficacy, a one unit increase leads to a .0111 increase in the first canonical variate of the *v-variables* set, when all of the other variables are held constant. At the bottom of the tables canonical correlation coefficients ($R_c$) are reported for each function. The strength of the relationship between the pairs of variates is reflected by the CCA coefficient ($R_c$). For the first function, $R_c = 0.8020$. For the second function, $R_c = 0.5798$. For the third function, $R_c = 0.1073$.

**Assessing Overall Model Fit**

The next part of Stata's output includes the multivariate tests for each function (see Figure 4). First, Wilk's lamda and corresponding *F*-tests, evaluate the null hypothesis that canonical correlations coefficients for all functions are zero.

Next, each function is evaluated against a null hypothesis that its canonical correlation coefficient is zero (i.e., the significance tests for canonical correlations 1, 2, 3). For this model, the first two canonical correlation coefficients are statistically significant (i.e., the null that the canonical correlation for a function equals zero is rejected or cannot be retained). The third function is not significant based on Wilk's lamda and corresponding F-tests, and will not be interpreted.

Figure 4

Tests of significance of all canonical correlations

| | Statistic | df1 | df2 | F | Prob>F | |
|---|---|---|---|---|---|---|
| Wilks' lambda | .234084 | 9 | 715.669 | 64.8885 | 0.0000 | a |
| Pillai's trace | .990959 | 9 | 888 | 48.6673 | 0.0000 | a |
| Lawley-Hotelling trace | 2.32123 | 9 | 878 | 75.4830 | 0.0000 | a |
| Roy's largest root | 1.80314 | 3 | 296 | 177.9098 | 0.0000 | u |

Test of significance of canonical correlations 1

| | Statistic | df1 | df2 | F | Prob>F | |
|---|---|---|---|---|---|---|
| Wilks' lambda | .234084 | 9 | 715.669 | 64.8885 | 0.0000 | a |

Test of significance of canonical correlations 2

| | Statistic | df1 | df2 | F | Prob>F | |
|---|---|---|---|---|---|---|
| Wilks' lambda | .656171 | 4 | 590 | 34.5890 | 0.0000 | e |

Test of significance of canonical correlation 3

| | Statistic | df1 | df2 | F | Prob>F | |
|---|---|---|---|---|---|---|
| Wilks' lambda | .988482 | 1 | 296 | 3.4491 | 0.0643 | e |

e = exact, a = approximate, u = upper bound on F

## Interpreting the Canonical Variates

**2.** ***canon, stdcoef***

When the variables in the model have different standard deviations, as they are in this example (see Figure 5), the standardized coefficients allow for easier comparisons among the variables. Since canonical correlation coefficients are standardized, their weights may be compared. The ratio of canonical correlation weights for a set of canonical roots is their relative importance for the given effect.

Figure 5

| Variable | SD |
|---|---|
| Parents' perceptions of self-efficacy | 5.31 |
| Relationship satisfaction between parents | 4.85 |
| Attitudes toward adoption | 4.90 |
| Service utilization | .50 |
| Satisfaction with parenting | 3.83 |
| satisfaction with adoption agency | 4.83 |

The standardized canonical coefficients for the first two (significant) functions are displayed in Figure 6. For the first variable set, *attitudes toward adoption* is most important, followed by *parents' perceptions of self-efficacy,* and *relationship satisfaction between parents.* The standardized canonical coefficients are interpreted in a manner analogous to interpreting standardized regression coefficients. For example, a one standard deviation increase in *parents' perceptions of self-efficacy* leads to a .0593 standard deviation increase in the score on the first canonical variate in the second variable set when the other variables in the model are held constant.

Figure 6

```
Canonical correlation analysis                    Number of obs =      300

Standardized coefficients for the first variable set

             |        1          2          3
-------------+---------------------------------
self_effic~y |  -0.0593    -0.4612     1.0162
relationsh~t |  -0.0060    -0.7062    -0.8714
attitude_a~n |   0.9993     0.0153     0.0981
-------------+---------------------------------

Standardized coefficients for the second variable set

             |        1          2          3
-------------+---------------------------------
service_ut~n |  -0.0374     0.9956    -0.1630
satisfacti~g |   0.9981     0.0465     0.0449
satisfacti~y |  -0.0280     0.0238     1.0090
-------------+---------------------------------
```

**3.** ***estat correlations***

It may be useful to list all correlations within and between sets of variables. Figure 7 displays (1) the within-set correlations for the *u*-variables (list 1) and the *v*-variables (list 2), and (2) the correlations between the *u*-variables and the *v*-variables.

Figure 7

Correlations for variable list 1

| | self_e~y | relati~t | attitu~n |
|---|---|---|---|
| self_effic~y | 1.0000 | | |
| relationsh~t | 0.4456 | 1.0000 | |
| attitude_a~n | 0.0121 | 0.0867 | 1.0000 |

Correlations for variable list 2

| | servic~n | satisf~g | satisf~y |
|---|---|---|---|
| service_ut~n | 1.0000 | | |
| satisfacti~g | -0.0042 | 1.0000 | |
| satisfacti~y | 0.1371 | -0.0187 | 1.0000 |

Correlations between variable lists 1 and 2

| | self_e~y | relati~t | attitu~n |
|---|---|---|---|
| service_ut~n | -0.4491 | -0.5281 | -0.0663 |
| satisfacti~g | -0.0570 | 0.0201 | 0.7983 |
| satisfacti~y | -0.0031 | -0.1298 | -0.0425 |

These univariate correlations must be interpreted with caution, since they do not indicate how the original variables contribute jointly to the canonical analysis. But, within and between-set correlations can be useful in the interpretation of the canonical variables. As displayed in Figure 2.7, for the *u*-variables, *relationship satisfaction between parents* is moderately associated with *parents' perceptions of self-efficacy* ($r$ = 0.4456). For the *v* variables, *satisfaction with the adoption agency* and *service utilization* are weakly correlated ($r$ = 0.1371). Consequently, these results can be interpreted to mean that *relationship satisfaction between parents* and *parents' perceptions of self-efficacy* are moderately and negatively related to *satisfaction with the adoption agency* and *service utilization* ($r$ = -0.4491 ) and ($r$ = -0. 5281) respectively. Moreover, attitudes toward adoption is strongly and positively associated with satisfaction with adoption agency ($r$ = 0.7983). That is, for this sample, it seems that adoptive parents with lower relationship satisfaction and perceptions of self-efficacy are more likely to be satisfied with the adoption agency. In addition, adoptive parents with positive attitudes to toward adoption are more likely to be satisfied with the adaption agency.

### 4. *estat loadings*

Next, the canonical loadings, sometimes termed structure coefficients, are displayed. These loadings are correlations between variables and the canonical variates (see Figure 8).

Figure 8

Canonical loadings for variable list 1

| | 1 | 2 | 3 |
|---|---|---|---|
| self_effic~y | -0.0499 | -0.7758 | 0.6290 |
| relationsh~t | 0.0542 | -0.9104 | -0.4101 |
| attitude_a~n | 0.9981 | -0.0515 | 0.0348 |

Canonical loadings for variable list 2

| | 1 | 2 | 3 |
|---|---|---|---|
| service_ut~n | -0.0454 | 0.9987 | -0.0249 |
| satisfacti~g | 0.9988 | 0.0419 | 0.0267 |
| satisfacti~y | -0.0518 | 0.1594 | 0.9858 |

For the $u$-variables, *attitudes toward adoption* is most closely related to the first canonical function, and *relationship satisfaction between parents* is most closely related to the second canonical function. For the $v$-variables, *satisfaction with the adoption agency* is most closely related the first canonical function, and *service utilization* is most closely related to the second canonical function.

### 5. *canred 1 /* findit canred */*

***6. canred 2 /* findit canred */***

Perform a canonical redundancy analysis.

Figure 9

```
Canonical redundancy analysis for canonical correlation 1

Canonical correlation coefficient          0.8020
Squared canonical correlation coefficient  0.6433

                                             own    opposite
Proportion of standardized variance  variate   variate
           of u variables with ...     0.3339    0.2148
           of v variables with ...     0.3341    0.2149
```

```
Canonical redundancy analysis for canonical correlation 2

Canonical correlation coefficient          0.5798
Squared canonical correlation coefficient  0.3362

                                             own    opposite
Proportion of standardized variance  variate   variate
           of u variables with ...     0.4778    0.1606
           of v variables with ...     0.3415    0.1148
```

$R_d$ is the amount of variance in a canonical variate explained by the other canonical variate in a canonical function. For example, for the first canonical function, the $R_d$ for the $u$-variables equals 0.2149, and the $R_d$ for the $v$-variables equals 0.2148. These values for each $R_d$ suggest that each canonical variate explains about the same amount of variance in the opposite set of variables in the first function. For the second canonical function, the $R_d$ for the $u$-variables equals 0.1148, and the $R_d$ for the $v$-variables equals 0.1606. These values for each $R_d$ suggest that the canonical variate for the u-variables explains more variance in the v-variables in the first function than the canonical variate for the $v$-variables explains in the set of $u$-variables (see Figure 9).

**Rotation of Structure and Canonical Coefficients**

Perform orthogonal Varimax rotation. A comparison of rotated and unrotated structure and canonical coefficients suggest that both solutions are equivalent. Equivalence between rotated and unrotated solutions suggests that these data have a simple structure and that this structure has been identified by the current CCA.

The rotated canonical correlations are usually expected to yield a more even distribution of variance among the canonical variates. This redistribution of variance usually results in a rotated structure that is more easily interpretable. This is analogous to changed distribution of factor variance contributions in factor analysis, following Varimax rotation. This is not the case with these data, and the equivalence between unrotated and rotated solutions suggests that the amount of predictable variance was not affected by the rotation. Since maximization is only present in the unrotated solution, the unrotated solution should be the focus of the description of results (see Figures 10 and 11).

**7.** ***estat rotate, stdcoefs***

*8.* ***estat rotate, loadings***

Figure 10

Rotated standardized coefficients

| | 1 | 2 | 3 |
|---|---|---|---|
| self_effic~y | 0.5057 | -0.0224 | 0.9963 |
| relationsh~t | 0.6663 | -0.0505 | -0.9009 |
| attitude_a~n | -0.0011 | 1.0026 | 0.0566 |
| service_ut~n | -1.0022 | -0.0325 | -0.1169 |
| satisfacti~g | -0.0347 | 0.9996 | 0.0049 |
| satisfacti~y | 0.0212 | 0.0143 | 1.0094 |

Figure 11

Rotated canonical loadings

| | 1 | 2 | 3 |
|---|---|---|---|
| self_effic~y | 0.8027 | -0.0327 | 0.5955 |
| relationsh~t | 0.8916 | 0.0265 | -0.4520 |
| attitude_a~n | 0.0628 | 0.9980 | -0.0095 |
| service_ut~n | -0.9992 | -0.0347 | 0.0214 |
| satisfacti~g | -0.0308 | 0.9994 | -0.0135 |
| satisfacti~y | -0.1156 | -0.0089 | 0.9933 |

## Reporting the Results of a CCA

The approach recommended here is to include a description of (1) variables n the model; (2) overall or omnibus hypotheses (*F*-value and *p*-value); (3) canonical correlation coefficient and canonical correlation coefficient square; (4) redundancy index; (standardized canonical coefficients; (5) canonical loadings or structure coefficients; and (6) rotated an unrotated solutions. Consolidating results into two tables or figures, one for the unrotated and one for the rotated solution, may help researchers who are new to this procedure interpret findings (see Figures 12 and 13 and Tables 1 and 2).

## Results of the Annotated Example

This study tested a model that links (1) factors influencing the utilization of post-adoption services (*parents' perceptions of self-efficacy*, *relationship satisfaction between parents*, and *attitudes toward adoption*) with (2) *service utilization* (two groups, used versus did not use post-adoption services), and (3) *positive adoption outcomes* (satisfaction with parenting and satisfaction with adoption agency).

Wilk's lamda and corresponding *F*-tests, were used to evaluate the null hypothesis that canonical correlations coefficients for all functions are zero. For this model, the first two canonical correlation coefficients are statistically significant, $p < .05$. The third function is not significant, and will not be interpreted.

The strength of the relationship between the pairs of variates is reflected by the CCA coefficient ($R_c$). For the first function, $R_c = 0.8020$. For the second function, $R_c = 0.5798$. For the second function, $R_c = 0.1073$. When squared, the canonical correlation represents the amount of variance in one optimally weighted canonical variate accounted for by the other optimally weighted canonical variate.

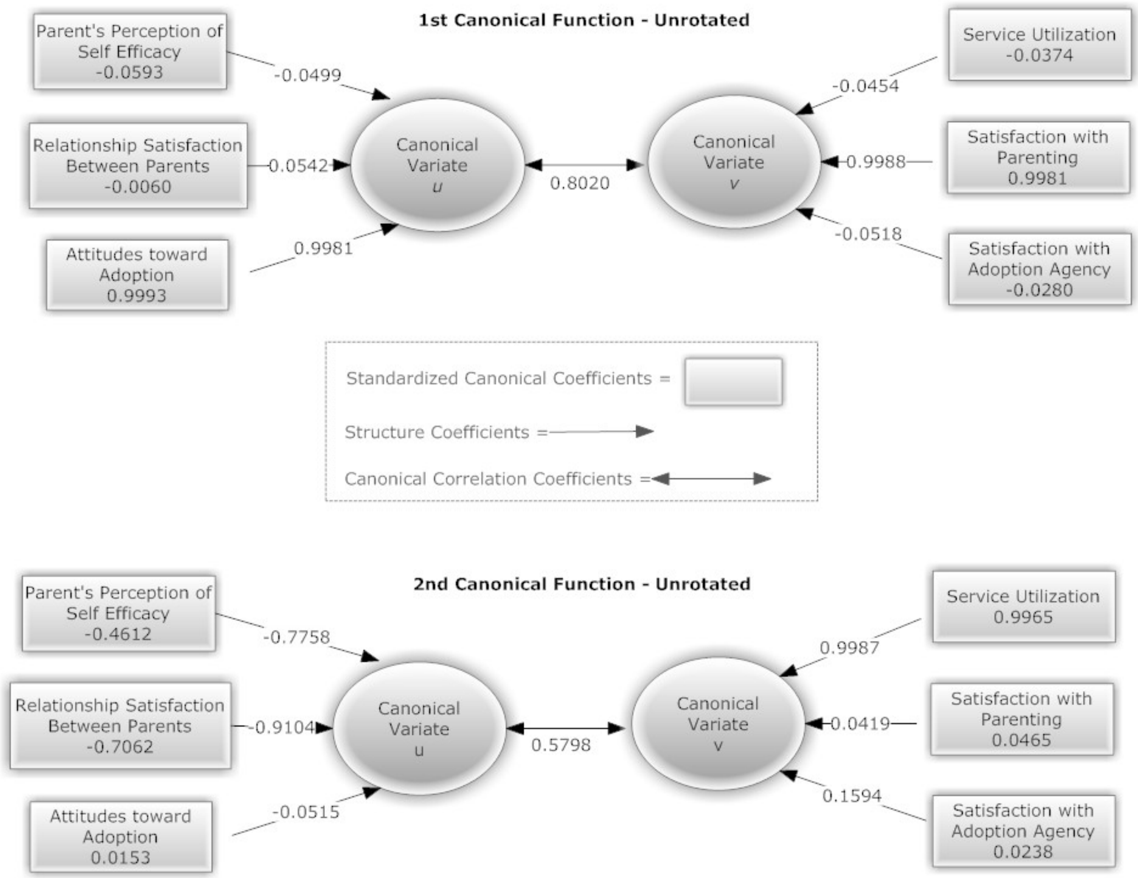
The redundancy index is a measure of the variance of one set of variables predicted from the linear combination of the other set of variables. The $\mathrm{R_d}$ is analogous to the squared multiple R in multiple regression. Recall that the redundancy coefficient can only equal 1 when the synthetic variables for the function represent all the variance of every variable in the set, and the squared $R_c$ also exactly equals 1. The redundancy index may be considered as a check on the meaning of the canonical correlation. For the first function $R_d = 0.2148$ for the *u*-variables, and $R_d = 0.2149$ for the *v*-variables. For the second function, $R_d = 0.1606$ for the *u*-variables, and $R_d = 0.1148$ for the *v*-variables.

Standardized canonical coefficients and canonical loadings were used to evaluate the relative importance of variables in the model. For the first variable set, *attitudes toward adoption* is most important, followed by *parents' perceptions of self-efficacy,* and *relationship satisfaction between parents.* The standardized canonical coefficients are interpreted in a manner analogous to interpreting standardized regression coefficients. For example, a one standard deviation increase in *parents' perceptions of self-efficacy* leads to a .0593 standard deviation increase in the score on the first canonical variate in the second variable set when the other variables in the model are held constant.

Canonical loadings are displayed in Figure 8. For the *u*-variables, *attitudes toward adoption* is most closely related to the first canonical function, and *relationship satisfaction between parents* is most closely related to the second canonical function. For the *v*-variables, *satisfaction with the adoption agency* is most closely related the first canonical function, and *service utilization* is most closely related to the second canonical function.

A comparison of rotated and unrotated structure and canonical coefficients implies that both solutions are equivalent (see Figures 12 and 13 and Tables 1 and 2). Equivalence between rotated and unrotated solutions suggests that these data have a simple structure and that this structure has been identified by the current CCA. Rotated canonical correlations are usually expected to yield a more even distribution of variance among the canonical variates. This redistribution of variance usually results in a rotated structure that is more easily interpretable. This is not the case with these data. Since maximization only is present in the unrotated solution, this solution should be the focus of the description of results.

Figure 12



1st Canonical Function - Unrotated
Parent's Perception of Self Efficacy -0.0593
-0.0499
Relationship Satisfaction Between Parents -0.0060
0.0542
Attitudes toward Adoption 0.9993
0.9981
Canonical Variate u
0.8020
Canonical Variate v
Service Utilization -0.0374
-0.0454
Satisfaction with Parenting 0.9981
0.9988
Satisfaction with Adoption Agency -0.0280
-0.0518
Standardized Canonical Coefficients =
Structure Coefficients =
Canonical Correlation Coefficients =
2nd Canonical Function - Unrotated
Parent's Perception of Self Efficacy -0.4612
-0.7758
Relationship Satisfaction Between Parents -0.7062
-0.9104
Attitudes toward Adoption 0.0153
-0.0515
Canonical Variate u
0.5798
Canonical Variate v
Service Utilization 0.9965
0.9987
Satisfaction with Parenting 0.0465
0.0419
Satisfaction with Adoption Agency 0.0238
0.1594

Figure 13

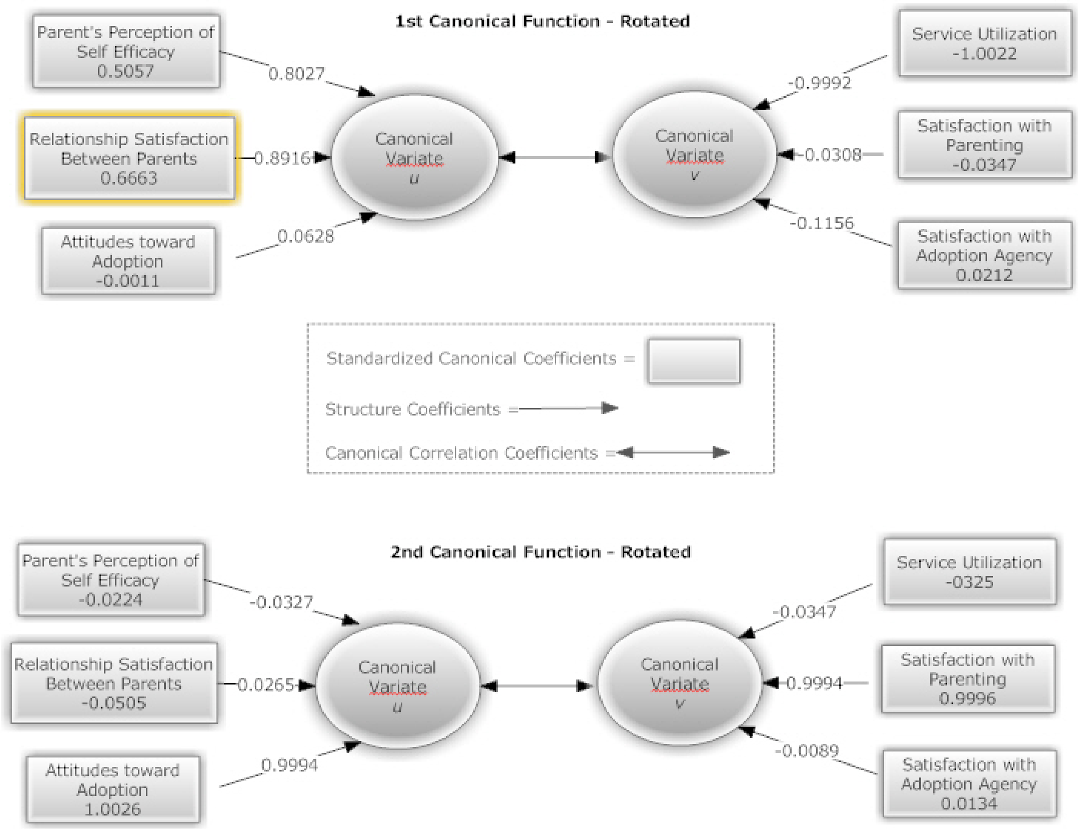



Table 1 Unrotated Solution

| | First Canonical Variate | | Second Canonical Variate | |
|---|---|---|---|---|
| ***Set 1*** | ***Coefficient***[a] | ***Loading***[b] | ***Coefficient*** | ***Loading*** |
| Parent's Perceptions of self-efficacy | -0.593 | -0.0499 | -0.4612 | -0.7758 |
| Relationship satisfaction between parents | -0.0060 | 0.0542 | -0.7062 | -0.9104 |
| Attitudes toward adoption | 0.9993 | 0.9981 | 0.0981 | -0.0515 |
| | | | | |
| ***Percent of Variance***[c] | 33.39 | | 47.88 | |
| ***Redundancy***[d] | 21.48 | | 16.06 | |
| | | | | |
| ***Set 2*** | ***Coefficient***[a] | ***Loading***[b] | ***Coefficient*** | ***Loading*** |
| Service utilization | -0.0374 | -0.0454 | 0.9956 | -0.0249 |
| Satisfaction with parenting | 0.9981 | 0.9988 | 0.0465 | 0.0267 |
| Satisfaction with adoption agency | -0.0280 | -0.0518 | 0.0238 | 0.9858 |
| | | | | |
| ***Percent of Variance*** | 33.41 | | 34.15 | |
| ***Redundancy*** | 21.49 | | 11.48 | |
| | | | | |
| ***Canonical Correlation***[e] | 0.8020 | | | |

[a]Standardized canonical variate coefficients
[b]Structure Coefficient
[c]Within-set variance accounted for by canonical variates (i.e., proportion of variance times 100)
[d]Percent of variance in one set of original variables explained by the other set's canonical variable
[e]Canonical correlations

Table 2 Rotated Solution (only coefficients and loadings are available in Stata)

| | First Canonical Variate | | Second Canonical Variate | |
|---|---|---|---|---|
| ***Set 1*** | ***Coefficient***[a] | ***Loading***[b] | ***Coefficient*** | ***Loading*** |
| Parent's Perceptions of self-efficacy | 0.5057 | 0.8027 | -1.00022 | -0.9992 |
| Relationship satisfaction between parents | 06663 | 0.8916 | -0.0347 | 0.0308 |
| Attitudes toward adoption | -0.0011 | 0.0628 | 0.0212 | -0.1156 |
| | | | | |
| ***Set 2*** | ***Coefficient***[a] | ***Loading***[b] | ***Coefficient*** | ***Loading*** |
| Service utilization | -0.0224 | -0.0327 | -0.0325 | -0.0347 |
| Satisfaction with parenting | -0.0505 | 0.0265 | 0.9996 | 0.9994 |
| Satisfaction with adoption agency | 1.0026 | 0.9994 | 0.0134 | 0.0089 |
| | | | | |

[a]Standardized canonical variate coefficients
[b]Structure Coefficient

## Conclusions

CCA is a useful and powerful technique for exploring the relationships among multiple dependent and independent variables. The technique is primarily descriptive, although it may be used for predictive purposes. This paper provided a demonstration of canonical correlation analysis with orthogonal rotation to facilitate interpretation. Results obtained from a canonical analysis can suggest answers to questions concerning the number of ways in which the two sets of multiple variables are related, the strengths of the relationships, and the nature of the relationships defined.

CCA enables the researcher to combine into a composite measure what otherwise might be an unmanageably large number of bivariate correlations between sets of variables. It is useful for identifying overall relationships between multiple independent and dependent variables, particularly when the data researcher has little a priori knowledge about relationships among the sets of variables. Essentially, the researcher can apply canonical correlation analysis to a set of variables, select those variables (both independent and dependent) that appear to be significantly related, and run subsequent analyses.

## References

Bagozzi, R. P., Fornell, C., & Larcker, D. F. (1981). Canonical correlation analysis as a special case of a structural relations model. *Multivariate Behavioral Research, 16*, 437-454.

Baggaley, A. R. (1981). Multivariate analysis: An introduction for consumers of behavioral research. *Evaluation Review, 5*, 123-131.

Barcikowski, R., & J. P. Stevens. (1975). A Monte Carlo study of the stability of canonical correlations, canonical weights, and canonical variate-variable correlations. *Multivariate Behavioral Research*, 10, 353-364.

Bartlett, M. S. (1948). Internal and external factor analysis. *British Journal of Psychology*, *1*, 73-81.

Carroll, J. D., & Green, P. E. (1997). *Mathematical tools for applied multivariate analysis*. New York: Academic Press.

Cleroux, R., & Lazraq, A. (2002). Testing the signicance of the successive components in redundancy analysis. *Psychometrika, 67*, 411-419.

Cliff*, N., &* Krus, J. D. *(*1976*).* Interpretation of canonical analysis: Rotated vs. unrotated solutions. *Psychometrika , 41, 35–42.*

Cooley, W. W., & Lohnes, P. R. (1971). *Evaluation research in education*. New York: Irvington.

Cramer, E. M., & Nicewander, W. A. (1979). Some symmetric, invariant measures of multivariate association. *Psychometrika*, *44*, 43–54.

Fan, X. (1997). Canonical correlation analysis and structural equation modeling: What do they have in common? *Structural Equation Modeling, 4*, 65-79.

Finn, J. (1978). *Multivariance: Univariate and multivariance analysis of variance, covariance and regression*. Chicago: National Education Resources.

Hair, J. F.., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis.* New York: Prentice Hall.

Hotelling, H. (1935). The most predictable criterion. *Journal of Educational Psychology*, *26*, 139-142.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, *28*, 321-377.

Knapp, T. R. (1978). Canonical correlation analysis: A general parametric significance testing system. *Psychological Bulletin*, *85*(2), 410-416.

Rencher, A. C. (1992). Interpretation of canonical discriminant functions, canonical variates, and principal components. *The American Statistician*, *46*, 3, 217-225.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences*. Hillsdale, NJ: Erlbaum.

Takane, Y., & Hwang, H. (2005). On a test of dimensionality in redundancy analysis. *Psychometrika*, *70*(2), 271-281.

Thompson, B. (1991). A primer on the logic and use on canonical correlation analysis. *Measurement and Evaluation in Counseling and Development, 24*(2), 80-95.

Warner, R. M. (2008). *Applied statistics*. Los Angeles, CA: Sage Publications.